Priors and Posteriors in Bayesian Timing of Divergence Analyses : The Age of Butterflies Revisited

Chazot, Nicolas

2019-09

Chazot , N , Wahlberg , N , Lucci Freitas , A V , Mitter , C , Labandeira , C , Sohn , J-C , Sahoo , R K , Seraphim , N , de Jong , R & Heikkilä , M 2019 , ' Priors and Posteriors in Bayesian Timing of Divergence Analyses : The Age of Butterflies Revisited ' , Systematic Biology , vol. 68 , no. 5 , pp. 797-813 . https://doi.org/10.1093/sysbio/syz002

http://hdl.handle.net/10138/307517
https://doi.org/10.1093/sysbio/syz002

Running head

THE AGE OF BUTTERFLIES REVISITED (AND TESTED)

Title

Priors and Posteriors in Bayesian Timing of Divergence Analyses: the Age of Butterflies Revisited.

Authors

NICOLAS CHAZOT[1,2,3*], NIKLAS WAHLBERG[1], ANDRÉ VICTOR LUCCI FREITAS[4], CHARLES MITTER[5], CONRAD LABANDEIRA[5,6], JAE-CHEON SOHN[7], RANJIT KUMAR SAHOO[8], NOEMY SERAPHIM[9], RIENK DE JONG[10], MARIA HEIKKILÄ[11]

Affiliations

[1]*Department of Biology, Lunds Universitet, Sölvegatan 37, 223 62, Lund, Sweden.*
[2]*Gothenburg Global Biodiversity Centre, Box 461, 405 30 Gothenburg, Sweden.*
[3]*Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, 405 30 Gothenburg, Sweden.*
[4]*Departamento de Biologia Animal, Instituto de Biologia, Universidade Estadual de Campinas (UNICAMP), Cidade Universitária Zeferino Vaz, Caixa postal 6109, Barão Geraldo 13083-970, Campinas, SP, Brazil.*
[5]*Department of Entomology, University of Maryland, College Park, MD 20742, U.S.A.*
[6]*Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA; Department of Entomology and BEES Program, University of Maryland, College Park, MD 20741; and Key Lab of Insect Evolution and Environmental Change, School of Life Sciences, Capital Normal University, Beijing 100048, China*
[7] *Department of Science Education, Gongju National University of Education, Gongju, Chungnam 32553, Republic of Korea*
[8]*IISER-TVM Centre for Research and Education in Ecology and Evolution (ICREEE), School of Biology, Indian Institute of Science Education and Research, Thiruvananthapuram, Kerala 695 551, India.*
[9]*Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campus Campinas, CTI Renato Archer - Av. Comendador Aladino Selmi, s/n - Amarais, Campinas - SP, 13069-901, Brazil.*
[10] *Naturalis Biodiversity Center, Department of Entomology, PO Box 9517, 2300 RA Leiden, The Netherlands.*

[11]*Finnish Museum of Natural History LUOMUS, Zoology Unit, P.O. Box 17, FI-00014 University of Helsinki, Finland.*

Corresponding author (*):

Nicolas Chazot

Department of Biology, Lunds Universitet, Sölvegatan 37, 223 62, Lund, Sweden.

email: chazotn@gmail.com

Abstract

The need for robust estimates of times of divergence is essential for downstream analyses, yet assessing this robustness is still rare. We generated a time-calibrated genus-level phylogeny of butterflies (Papilionoidea), including 994 taxa, up to 10 gene fragments and an unprecedented set of 12 fossils and 10 host-plant node calibration points. We compared marginal priors and posterior distributions to assess the relative importance of the former on the latter. This approach revealed a strong influence of the set of priors on the root age but for most calibrated nodes posterior distributions shifted from the marginal prior, indicating significant information in the molecular dataset. Using a very conservative approach we estimated an origin of butterflies at 107.6 Ma, approximately equivalent to the latest Early Cretaceous, with a credibility interval ranging from 89.5 Ma (mid Late Cretaceous) to 129.5 Ma (mid Early Cretaceous). In addition, we tested the effects of changing fossil calibration priors, tree prior, different sets of calibrations and different sampling fractions but our estimate remained robust to these alternative assumptions. With 994 genera, this tree provides a comprehensive source of secondary calibrations for studies on butterflies.

Keywords

Papilionoidea, butterflies, time-calibration, fossils, host plants, marginal prior, Early Cretaceous

INTRODUCTION

An increasing amount of molecular information is allowing the inference of broad and densely sampled phylogenetic hypotheses for species-rich groups. This effort, combined with the emergence of a great number of methods investigating trait evolution, historical biogeography, and the dynamics of diversification have increased the need for time-calibrated trees. Estimating divergence times in molecular phylogenetic work depends primarily on fossils to constrain models of heterogeneous rates of substitutions. Consequently, the robustness of such estimates relies on the quality of fossil information, involving age and taxonomic assignment (Parham et al 2012), the priors assigned to nodes that are calibrated in a Bayesian analysis (Warnock et al 2012, Brown & Smith 2017), and the amount of information inherent in the molecular dataset (Yang & Rannala 2006, Rannala & Yang 2007, dos Reis & Yang 2013).

Fossils inform us of the minimum age of a divergence, imposing a temporal constraint that is widely accepted. However, the constraint of a simple hard minimum age is insufficient information for a proper analysis of times of divergence, particularly as there is an absence of information about maximum ages for divergences, including the root node. Often fossil information is modeled as a probability distribution, such as a lognormal or exponential distribution, indicating our beliefs regarding how informative a fossil is about the age of a divergence (Drummond et al 2006, Warnock et al 2015). The distributional shapes of these priors are often established without justification (Warnock et al 2012). Ideally, in node-based dating, fossil information is used only as a minimum age constraint for a given divergence in the form of a uniform prior with a minimum age equaling the fossil age and a maximum age

extending beyond the age of the clade in question. In such cases at least one maximum constraint is needed, often also based on fossil information. Another approach is use of additional information, such as using ages of host-plant families as maximum constraints for highly specialized phytophagous insect clades (Wahlberg et al 2009). In such cases, a uniform prior also can be used, with the maximum set to the age of the divergence of the host-plant family from its sister group and the minimum set to the present time.

Brown & Smith (2017) recently have pointed out the importance of assessing the relative influence of priors over the actual amount of information contained in the molecular dataset. As noted above, users specify fossil calibrations using prior distributions by modeling the prior expectation about the age of the node constrained. However, the broader set of fossil constraints can interact with each other and with the tree prior, leading to marginal prior distributions at nodes that usually differ from the user's first intention (Warnock et al 2012). If relevant information were contained within the molecular dataset, one would expect the posterior distribution to shift from the marginal prior distribution. In the case of angiospermous plants, Brown & Smith (2017) showed that the marginal prior resulting from the interaction of all priors (fossils and the tree) excluded an Early Cretaceous origin, in effect giving such an origin zero probability. In addition, many calibrated internal nodes showed nearly complete overlap of marginal prior and posterior distributions, suggesting little information in the molecular dataset but a potentially strong influence of the set of priors.

With more than 18,000 species described and extraordinary efforts made to infer phylogenetic hypotheses based on molecular data, butterflies (Lepidoptera:

Papilionoidea) have become a model system for insect diversification studies. Nevertheless, the paucity of information available to infer times of divergence in butterflies questions the reliability of the various estimates (e.g. Garzón-Orduña et al. 2015). Heikkilä et al (2012) for example, used only three fossils to calibrate a deep-level phylogeny of the superfamily Papilionoidea. The shortage of fossil information for calibrating large-scale phylogenies also means that, most of the time, species-level phylogenies at a smaller scale rely on secondary calibration points extracted from the deep-level time-trees (e.g. Peña et al 2011, Matos-Maravi et al 2013, Kozak et al 2015, Chazot et al 2016, Toussaint & Balke 2016).

In a recent paper, de Jong (2017) revisited the butterfly fossil record, providing a discussion about the quality of the different fossil specimens as well as their taxonomic placement. Using this information, we established an unprecedented set of 12 fossil calibration points across all butterflies, which we use in this study to revisit the timescale of butterfly evolution in a comprehensive phylogenetic framework, and investigate the robustness of this new estimate. We complement the minimum age constraints of clades based on fossils with maximum age constraints based on the ages of host-plant families. Some clades of butterflies have specialized on specific groups of angiosperm hosts for larval development, such that one may assume that diversification of the associated butterfly clade only occurred after the appearance of the host-plant clade. We use this assumption as additional information to calibrate the molecular clock by setting the age of specific clades of butterflies to be younger than the estimated age of their host-plant lineage. We restrained these calibrations to deep-level host-plant clades.

The first studies of divergence times using representatives of all butterfly families inferred a crown clade age of butterflies of 110 Ma (Heikkilä et al 2012) and 104 Ma (Wahlberg et al 2013), which implied a large gap from the oldest known fossil of butterfly, estimated to be 55.6 Ma and confidently assigned to the extant family Hesperiidae (de Jong 2016, 2017). Such discrepancy has been extensively debated for a similar case, the origin of angiosperms, often estimated to have originated during the Triassic (252–201 Ma ago), while the oldest undisputed fossil is pollen dated at 136 Ma. Despite a much more fragmentary fossil record for butterflies, the same questions remain. First, are the previous estimates robust to a more comprehensive assemblage of fossils and taxon sampling? Second, is the 52 million-year discrepancy between molecular clock estimates and the fossil record accurate or the result of a lack of information contained in the molecular dataset? In other words, how much does the set of priors influence the results?

Here, we generated a genus-level phylogeny of Papilionoidea, including 994 taxa, in order to maximize the number and position of fossil calibration points and increase the potential amount of molecular information. By establishing the set of 12 fossils and 10 host-plant calibration points, we time-calibrated the tree in order to provide a revised estimate of the timing in diversification of butterflies. We then assessed the robustness of these results to the assumptions made throughout the analysis, including (i) different subsets of time constraints, (ii) the prior distributions of fossil constraints, (iii) a different estimate for host-plant ages, (iv) a Yule tree prior, (v) a reduced taxon sampling and (vi) the addition of a mitochondrial gene fragment to the nine nuclear gene regions.

Finally, we compared the user specified priors, marginal prior and posterior distributions of different analyses, to assess the influence of our set of constraints on the estimated timing of divergences.

MATERIALS AND METHODS

Molecular Dataset

When designing our dataset, we aimed at building a genus-level tree of Papilionoidea. We assembled a dataset of 994 taxa from the database VoSeq (http://www.nymphalidae.net/db.php, Peña & Malm 2012), with each taxon representing a genus. Overall, ~54% of butterfly genera were included in our tree (Papilionidae: 100%, Hedylidae: 100%, Hesperiidae: ~50%, Pieridae: ~97%, Lycaenidae: ~14%, Riodinidae: ~62%, Nymphalidae: ~88%). We chose to include gene fragments that were available across the whole tree in order to avoid large clade-specific gaps in the molecular dataset. In addition, Sahoo et al (2016) pointed out a conflicting signal in the family Hesperiidae between nuclear and mitochondrial markers. Thus, we chose to primarily focus on nuclear markers. Our final dataset included nine gene fragments: ArgKin (596bp), CAD (850bp), EFI-α (1240 bp), GAPDH (691bp), IDH (710 bp), MDH (733 bp), RPS2 (411 bp), RPS5 (617 bp) and wingless (412 bp) for a total length of 6260 base pairs. The list of taxa, Genbank accession codes and data matrix are available in the Supplementary Material S1-2 (Dryad Digital Repository: doi:10.5061/dryad.fb88292).

Set of Time-Calibrations for Timing Analyses

178 Fossil calibrations – Previous studies estimating times of divergence of butterfly
179 lineages have largely relied on unverified fossil calibrations. The identifications of
180 these calibrations were often based on overall similarity with extant taxa, not
181 apomorphies. In the present study, we initially chose 14 fossil butterflies that were
182 recently critically reviewed by de Jong (2017) and displayed apomorphic characters
183 or character combinations diagnostic of extant clades, thereby allowing reliable
184 allocation of fossils on the phylogenetic tree to provide minimum ages to the
185 corresponding nodes. These fossils included three inclusions in Dominican Amber
186 and 11 compression/impression fossils. For the age of these fossils we have relied on
187 the most recent dates established from recent advances in Cenozoic
188 chronostratigraphy, geochronology, chemostratigraphy and the geomagnetic polarity
189 time scale (Walker et al 2013). These improvements by geologists and specialists in
190 allied disciplines have provided an increased precision in age dates of stratigraphic
191 record (International Commission on Stratigraphy, 2012). The list of fossils and their
192 positions in the tree is given in Table 1 and in Supplementary Material S12. For more
193 detailed information on the identification of these fossils, localities, preservation type
194 and current depositories, see de Jong (2017).

195 When a fossil was assigned to a clade, we calibrated the stem age of this clade,
196 specifically the time of divergence from its sister clade, instead of the crown age or
197 the first divergence event recorded in the clade of interest. As a consequence of this
198 choice, we removed two of the 14 fossils. We did not use *Praepapilio colorado*
199 Durden & Rose, 1978 (Papilionidae, 48.4 Ma) nor the less well-preserved *Praepapilio*
200 *gracilis* Durden & Rose, 1978 (Papilionidae) of the same age because its position at
201 the root of the tree was uninformative given the presence of the 55.6 million years old
202 *Protocoeliades kristenseni* de Jong, 2016 placed at the crown of the Hesperiidae. For

similar reasons, we did not use *Doxocopa wilmattae* Cockerell, 1907 (Nymphalinae+Biblidinae+Limenitidinae+Apaturinae, 33.8 Ma) because its position was uninformative given the presence of *Vanessa amerindica* Miller & Brown, 1989 of the same age but placed lower in the tree.

Host-plant calibrations – Butterflies are well known for their strict relationships with specific groups of plants used by their larvae. Such associations have previously been suggested as evidence for coevolution (Ehrlich & Raven, 1964, Janz & Nylin 1998, Nylin & Janz 1999). In the present study, we selected nine calibration points based on known information of host-plant specificity by butterflies since the large revision of Ackery (1988) (see also Beccaloni et al 2008 for Neotropical species), and revised for those host-plant records listed as having spurious or occasional records (André V.L. Freitas unpublished data). Host-plant clades used by single genera or a small group of recently-derived genera were discarded, such as the use of Aristolochiaceae by Troidini. In these cases the butterflies clearly are much more recent than their associated plant clades, and consequently do not contribute relevant time information to the tree. We defined the ages of each plant group as maximum ages for the respective nodes. For all host-plant maximum constraints we used the estimate from Magallón et al (2015) using the upper boundary of the 95% credibility interval of the stem age of the host-plant clade. We also constrained the root of the Papilionoidea with a maximum age corresponding to the crown age of angiosperms from Magallón et al (2015). The host-plant calibrations were placed at the crown of the butterfly clades as a conservative approach since we do not know when the host-plant shift occurred on the stem branch. However, we assume that the diversification of the clade could not have begun earlier than the origin of the host-plant family. The list of host

plant calibration points and their positions in the tree is given in Table 1 and in Supplementary Material S12.

Analyses Overview

Given computational limitations for such a dataset, we adopted the following procedure (details given below). We ran PartitionFinder v. 1.1 (Lanfear et al 2012) to identify the best partition scheme. Using this result, we performed a maximum likelihood analysis to obtain a tree topology. This tree topology was transformed into a time-calibrated ultrametric tree and used thereafter as a fixed topology and starting tree in all our dating analyses. Branch lengths were estimated using BEAST v. 1.8.3 (Drummond et al 2012) with a simpler partitioning scheme, a birth-death tree prior, lognormal relaxed molecular clocks, and a combination of minimum (fossils) and maximum (host-plants) constraints for which all were set with uniform priors. This constituted the core analysis. We then performed additional analyses to test the robustness of our results to (i) different subsets of time constraints, (ii) the prior distribution of fossil constraints, (iii) a different estimate for host-plant ages, (iv) a Yule tree prior, (v) a reduced taxon sampling, and (vi) the addition of a mitochondrial gene fragment.

Core Analysis

Tree topology – We started by running PartitionFinder v. 1.1 (Lanfear et al 2012) on the concatenated dataset, allowing all possible combinations of codon positions of all genes. Substitution models were restricted to a GTR+G model and branch lengths

were linked. We then performed a maximum likelihood analysis using RAxML v8 (Stamatakis 2006) using the best partitioning scheme identified by PartitionFinder and 1000 rapid bootstraps (Supplementary Material S3). The resulting tree was set as a fixed topology for the dating analyses. To do so, the tree was transformed into a time-calibrated ultrametric tree using the package *ape* (Paradis et al 2004) and the full set of minimum and maximum calibrated nodes in order to obtain a starting tree suitable for BEAST analyses.

Time tree – We used BEAST v. 1.8.3 (Drummond et al 2012) to perform our time-calibration analysis. Given the size of our dataset, we reduced the number of partitions in our dating analysis to three partitions, each partition being one codon position of all genes pooled together. Substitution rate for each partition was modeled by GTR+G and an uncorrelated lognormal relaxed molecular clock. We used a Birth-Death process as the branching process prior. In order to have a fixed topology we turned off the topology operators in BEAUTi and we specified the topology obtained with RAxML made ultrametric with the *ape* package.

Setting the priors for calibration points is always an important matter of discussion. Non-uniform priors are often used, yet in the majority of studies the choice of parameters defining the shape of the prior distribution is not justified (Warnock et al 2012). For the core analysis we followed a conservative approach – considering that fossils only provide a minimum age, while host-plant calibrations only provide a maximum age for the nodes they were assigned to – and we used uniform prior distributions for all calibration points (Table 1). When a node was calibrated with fossil information, the distribution ranged from the estimated age of the fossil to the age of angiosperm origin (extracted from Magallón et al 2015). When a node was

calibrated using host-plant age, the prior distribution ranged from 0 (present) to the age of the host-plant clade origin. When a node was calibrated with both types of information, the distribution ranged from the age of the fossil to the age of host-plant clade origin. We also used a uniform prior for the tree root height, ranging between the oldest fossil used in the analysis and the age of angiosperm origin. Host-plant calibrations, as well as the origin of angiosperms were extracted from Magallón et al (2015), using the upper boundary of the 95% credibility interval of the stem age of the host-plant clade. Our choice of combining (1) uniform prior distributions, (2) fossil calibration of stem nodes, (3) the oldest stem age of the host-plant clades and (4) host-plant calibration of crown nodes has important implications. On the one hand these choices are the most conservative options, cautiously using the information given by each type of calibration point and taking into account uncertainty surrounding the information used. On the other hand, they are also the least informative.

We performed four independent runs of 30 million generations, sampling every 30 000 generations. We checked for a satisfactory convergence of the different runs using Tracer v. 1.6.0 (Rambaut et al 2014) and the effective sample size values in combination. Additionally, we performed three independent runs of 70 million generations, sampling every 7 000 generations. Using Tracer v. 1.6.0 we compared posterior distributions of the short runs with the long runs. Both analyses were convergent, and we used 30 million generation runs for all subsequent analyses, unless stated otherwise. Using LogCombiner v. 1.8.3 (Drummond et al 2012), we combined the posterior distributions of trees from the three runs, discarding the first 10% of trees of each run. Using TreeAnnotator v. 1.8.3 (Drummond et al 2012) we extracted the median and the 95% credibility interval of the posterior distribution of node ages.

299

300 *Alternative Analyses*

301 We tested the effect of making alternative choices along the core analysis on our
302 estimates of divergence times. Unless stated otherwise, we made only one
303 modification at a time; all other parameters remained identical to that described for
304 the core analysis. We performed at least two independent runs of 30 million
305 generations per alternative parameter set and more if convergence was not reached.

306 Different subsets of fossils – We aimed at testing whether using only a
307 fraction of the fossil information affected the estimation of divergence times and
308 whether the position of calibrations (close to the root or close to the tips) also changed
309 the results. Thus, we divided our set of fossil constraints into two subsets depending
310 on their position in the tree. One subset included fossil calibration points assigned at a
311 deep level in tree (hereafter: deep-level fossils): *Lethe*, *Mylothrites*, *Neorinella*,
312 *Pamphilites*, *Prolibythea*, *Protocoeliades* and *Vanessa* (Table 1). The other subset
313 included fossil calibration points close to the tips of our phylogeny (hereafter:
314 shallow-level fossils): *Doritites*, *Thaites*, *Dynamine*, *Theope* and *Voltinia* (Table 1). In
315 both cases the full set of maximum constraints was used. We performed one analysis
316 for each subset.

317 Exponential fossil priors – In the core analysis we used uniform distributions
318 for calibration points, which is a conservative option but also the least informative. As
319 an alternative, we designed exponential priors for fossil calibration points.
320 Exponential priors use the age of a fossil as minimum age for the node it has been
321 assigned to, but also assume that the probability for the age of the node decreases
322 exponentially as time increases. In BEAUTi, we set the offset of exponential

distributions with the age of the fossil. The distribution was truncated at the maximum age used in the uniform priors. The shape of the exponential distribution is controlled by a mean parameter, which has to be arbitrarily chosen by the users. The choice of mean parameter can be found in Table 1. Priors for host-plant calibration points were not changed (i.e., uniform priors).

Yule branching process prior – Condamine et al (2015) showed that the prior for the tree growth can a have a great impact on the estimated divergence times. In the core analysis we used a Birth–Death prior, which models the tree formation with a constant rate of lineage speciation and a constant rate of lineage extinction. As an alternative, we used a Yule prior, which involved a constant rate of speciation and no extinction to assess whether age estimates changed or not.

Alternative host-plant ages –The origin and timing of diversification of angiosperms is controversial. While the oldest undisputed fossil of Angiospermae is from the mid Early Cretaceous (136 Ma, Brenner 1996), most divergence time estimations based on molecular clocks have inferred a much older origin. In the core analysis, we chose to use host-plant ages derived from the tree of angiosperms time-calibrated by Magallón et al (2015), who imposed a constraint on the origin of angiosperms based on this fossil information. They found a crown age for angiosperms of ~ 140 Ma. As an alternative consistent with an older origin of angiosperms we used ages recently inferred by Foster et al (2017), who recovered a crown age of angiosperms of ~ 209 Ma. All maximum constraints were replaced by those inferred by Foster et al (2017). The origin of angiosperms used as a maximum constraint was set to the upper boundary of the 95% credibility interval of the crown age of the angiosperms i.e., 252.8 Ma. Because the posterior distributions of node

ages for this analysis were very skewed, we extracted the median of the distribution, the 95 % credibility interval and the mode of the kernel density estimate of nodes using the R package *hdrcde*. For comparison, we also estimated the mode of posterior distributions for the core analysis and all alternative tests.

Using only fossil information – As another alternative set of constraints we performed an analysis using only fossil information (no maximum age based on host-plant information), however modeled using the more informative lognormal prior distributions. The shape of the distributions is designed by a mean parameter, a standard deviation and the offset, which are all defined arbitrarily by the users. The parameters used here can be found in Table 1. We performed two runs of 60 million generations for this analysis.

Reduced dataset – In our core analysis, we chose to maximize the taxon sampling – increasing the number of lineages – which increased the fraction of missing data in the molecular dataset. We tested whether increasing the molecular dataset completion to the detriment of taxon sampling changed the results. In this reduced dataset, we included all the genera for which a specific minimum number of genes were available. The missing data in the molecular dataset are not uniformly distributed across the tree; for example, Lycaenidae have more missing data than the Nymphalidae. Therefore, a different cut-off value was chosen for each family in order to keep a good representation of the major groups (Papilionidae: 5 genes, Hedylidae: 8 genes, Hesperiidae: 9 genes, Pieridae: 8 genes, Lycaenidae: 4 genes, Riodinidae: 8 genes, Nymphalidae: 9 genes). In order to allow assignment of all fossils to the same place as in the core analysis, nine taxa having a number of genes below the cut-off value had to be added. We ended up with a dataset reduced to only 364 taxa instead of

994 in the core analysis. Accordingly, the fraction of missing data decreased from 39.5% in the core analysis to 21.4% (Supplementary Material S2). Given this important modification of the dataset we generated a new topology with RAxML, which was then calibrated identically to the core analysis.

Mitochondrial gene fragment – We tested whether adding mitochondrial information in the dataset would affect our results. To do so, we added the *cytochrome oxidase subunit I* (COI) gene to the molecular dataset. Given the conflicting signal in Hesperiidae between nuclear and mitochondrial information (Sahoo et al. 2016), the COI was not added to the Hesperiidae (Supplementary Material S2). We performed a new RAxML analysis in order to obtain a new topology. This new tree was calibrated with BEAST identically to the core analysis, with one difference. The mitochondrial gene was added as two partitions separated from the nuclear partitions: the first and second positions of COI were pooled together and the third position had its own partition. Therefore this analysis had five partitions.

Comparing Prior and Posterior Distributions

When performing a Bayesian analysis, comparing prior and posterior parameter distributions can be informative about the amount of information contained by our data compared to the influence of prior information. As exemplified by Brown & Smith (2017), such a comparison can shed light on the discrepancies observed in the fossil record and the divergence times estimated from a time-calibrated molecular clock. It may also help to disentangle the effect of interaction among calibration points. For each calibrated node we can compare the user-designed prior distribution (e.g., uniform distributions in the case of the core analysis), the marginal prior

distribution that is the result of the interaction between the user priors and the tree prior, and the posterior distribution that is the distribution after observing the data.

For the core analysis, the two different subsets of fossils and the alternative host-plant ages analyses were re-run without any data to sample from the marginal prior. In each case we performed two independent runs of 50 million generations, sampling every 50 000 generations. The results were visualized with Tracer. When necessary, we performed an additional run. Using LogCombiner, the runs were combined after deleting the first 10% as burn-in. The results of the analyses with and without the molecular dataset were imported into R (R Development Core Team 2008) and for each calibrated node as well as the root height we compared the kernel density estimates of the marginal prior and the posterior distributions (R package *hdrcde*).

Comparison with Previous Studies

For the root of all Papilionoidea and the seven families we compared the estimates obtained in the core analysis to previous studies that also used fossil information.

RESULTS

Core Analysis

The core analysis performed with BEAST used the full set of fossils and host-plant constraints from Magallón et al. (2015) on the topology found with RAxML. This analysis resulted in a root estimate for all Papilionoidea of 107.6 Ma (Fig.1, Supplementary Material S3 and S12). The 95% credibility interval of the posterior distribution ranged from 88.5 to 129.5 Ma. The lineage leading to Papilionidae

417 diverged first at the root of Papilionoidea and the crown age of Papilionidae was
418 inferred to be 68.4 Ma (95%, CI=53.5–84.3). Hedylidae and Hesperiidae diverged
419 from Pieridae–Lycaenidae–Riodinidae–Nymphalidae at 106.5 Ma (95%, CI=88.0–
420 127.2) and diverged from each other at 99.2 Ma (95%, CI=80.7–119.2). The crown
421 age of the sampled Hedylidae was 32.8 Ma (95%, CI=23.4–43.6) and crown age of
422 Hesperiidae was 65.2 Ma (95%, CI=55.8–78.1). Pieridae diverged from Lycaenidae–
423 Riodinidae–Nymphalidae at 101.1 Ma (95%, CI=83.0–120.3) and extant lineages
424 started diversifying around 76.9 Ma (95%, CI=63.1–92.4). Lycaenidae and
425 Riodinidae diverged from Nymphalidae at 97.4 Ma (95%, CI=80.4–116.5) and
426 diverged from each other at 87.8 Ma (95%, CI=73.2–106.1). The crown age of
427 Lycaenidae was 71.0 Ma (95%, CI=57.2–85.2) and crown age of Riodinidae was 73.4
428 Ma (95%, CI=60.3–88.1). Finally, the crown age of Nymphalidae was inferred to be
429 82.0 Ma (95%, CI=68.1–98.3). The complete tree, including median node ages,
430 credibility intervals, and the positions of fossil and host-plant calibration points are
431 shown in Supplementary Material S12.

432

433 *Alternative Analyses*

434 In most cases the eight alternative parameters tested yielded very similar results (Fig.
435 2, Supplementary Material S4-S11). Reducing the number of taxa in order to decrease
436 the fraction of missing data, using deep-level calibration points only, or using a Yule
437 process tree prior (instead of a Birth–Death prior), gave virtually identical results as
438 the core analysis above. Using only shallow-level fossil constraints (close to the tips
439 of the phylogeny) resulted in the youngest estimates of all alternative runs, with a
440 crown age of Papilionoidea of 94.5 Ma (mode=83.8, 95%, CI=67.8–126.6). Using

exponential fossil priors mainly resulted in a narrower credibility interval, while the mode and median age estimates were only 7–8 million years younger than the core analysis mode estimate (Fig. 2, Supplementary Material S7). Adding mitochondrial information also lead to a 7–8 million-year younger estimate for the crown age of Papilionoidea, but the credibility interval remained comparable to the core analysis (Supplementary Material S8). Finally, using a hypothesis of older host-plant ages extracted from Foster et al (2017), we obtained the greatest difference. The upper boundary of the credibility interval largely shifted toward much older ages (95%CI=88.5–167.2) as well as the median (119.5 Ma). The posterior distribution was, however, very skewed, with a mode of 101.0 Ma, and converged to the same age as the core analysis (Fig. 2, Supplementary Material S10-11). When running analyses with only fossil information but lognormal priors we recovered estimates identical to the core analysis but with a narrower credibility interval.

These variations for the root age among different alternative analyses were also reflected in the estimated ages of the different families. For example, all shallow-level fossils always led to younger estimates while older ages from Foster et al. (2017) always led to older estimates (Fig. 2).

Comparing Prior and Posterior Distributions

We compared the posterior distributions to the marginal prior distributions for the different calibrated nodes in the core analysis. We set all fossil and host-plant constraints with uniform prior distributions as we considered this as the most conservative approach. However, it is important to note that the marginal prior distributions at these nodes, which result from the interactions between all calibration priors and tree prior, are not uniform (Fig. 3).

Across all calibrated node points, many of them showed shifts of posterior distributions from the marginal priors, indicating that the results of the core analysis were not a simple outcome of our set of priors (Fig. 3). Interestingly, the nodes calibrated by *Doritites*, *Dynamine*, *Thaites*, *Theope* and *Voltinia*, which are all the fossils placed close to the tips of our phylogeny, tended to shift away from the minimum boundary, toward older ages than the marginal prior distribution. Alternative analyses performed with only these shallow-level fossils yielded the youngest tree for butterflies. This suggests that deep-level fossils bring important additional information, leading posterior distributions of shallow-level nodes to shift away from the prior distributions in the core analysis.

The nodes calibrated with the deep-level fossils *Mylothrites*, *Prolibythea*, *Neorinella* and *Vanessa* showed posterior distributions largely overlapping with their marginal prior distributions. Many host-plant calibrated points showed a shift from the marginal prior distribution (Fig. 3). In all cases, except the node also calibrated with the fossil *Lethe*, the crown age of the butterfly clade inferred was much younger than the age of the corresponding host-plant clade.

For the root of Papilionoidea, the marginal prior and posterior distributions largely overlapped in the core analysis, therefore not indicating whether our molecular dataset contained significant information about the root age or not. We also compared the posterior and the marginal prior distributions for alternative analyses performed with different subsets of fossil calibrations (Fig. 4). When using only deep-level fossils, the posterior distribution was almost identical to the core analysis, but the marginal prior slightly shifted from the marginal prior of the core analysis toward a younger age. The use of only shallow-level fossils had more profound effects. In such

a case, prior distributions of the core analysis and the shallow-level fossil alternative completely overlapped. The posterior distribution, however, shifted toward younger ages, yielding the most recent estimate for the root age among all analyses (mean=94.5, mode=83.8, 95%, CI=67.8–126.5). We also looked at the effect of using relaxed maximum ages (based on Foster et al. 2017). In this case, marginal prior distribution for the root age shifted to a mean of ~148 Ma (Fig. 4) and a credibility interval spanning 100 Ma (95%, CI=99.9–205.8). The posterior distribution was very skewed, retaining a wider credibility interval than the core analysis (95%, CI=88.5–167.5), but significantly shifted from the prior distribution toward the posterior distribution of the core analysis (median=119.5, mode=101.0).

Comparison with Previous Studies

For the root of Papilionoidea, our estimate in the core analysis using the mode age of the distribution was very similar to Wahlberg et al (2013) and Heikkilä et al (2012), with a mean age estimate of 104.6 and 110.8 Ma, respectively (107.6 Ma in the core analysis, Fig. 5). Espeland et al (2018) using a reduced taxon sampling and set of time-calibrations but a large genomic dataset obtained similar time for the origin of butterflies of 118.3 (95%, CI=91.2–142.5) as well. In a recent mitogenomic time-calibrated tree, however, Condamine et al (2018) obtained contrasting results. When using a single molecular clock for their dataset they recovered similar ages as found here, yet with a large credibility interval (98.4, 95%, CI=66.16–188.58). When partitioning their dataset into 11 molecular clocks however, they found a mean time of origin about 30 million years younger (71.27, 95%, CI=64.25–86.2).

For the crown age of families our estimates were often consistent with most of previous studies. We note that all published studies have used very different sets of

513 calibrations, priors, taxon sampling and gene region sampling, all factors leading to
514 different estimates for ages. For Papilionidae, our crown age estimate (68.4, 95%,
515 CI=53.5–84.3) was very similar to Wahlberg et al (2013) and Heikkilä et al (2012)
516 and slightly younger than the two recent phylogenomic studies (Espeland et al 2018,
517 Condamine et al 2018). Condamine et al (2012), however, in a study focusing also on
518 Papilionidae found younger ages by about 15 million years. For Hedylidae, only
519 Heikkilä et al (2012) and Espeland et al (2018) had an estimate for the crown age,
520 about 10 million years older than our result (32.8, 95%, CI=23.4–43.6) for Heikkilä et
521 al (2012) but very similar for Espeland et al (2018). The mean crown ages for the
522 Hesperiidae published so far range from 58.31 Ma (Condamine et al 2018, one clock)
523 to 82 Ma (Sahoo et al 2017) and the estimate fell within this range (65.2, 95%,
524 CI=55.8–78.1 in our study. Pieridae is the family that showed greatest variation in age
525 estimates among different studies. Our estimate (76.9 Ma, 95%, CI=63.1–92.4 Ma)
526 falls between the youngest estimate from Wahlberg et al (2013) and the oldest
527 estimate from Braby et al (2006), for which credibility intervals did not overlap. Our
528 estimate was very similar to the recent phylogenomic study by Espeland et al (2018).
529 For Lycaenidae, which lack fossils calibrations, the results among our core analysis
530 (73.4, 95%, CI=60.3–88.1), Wahlberg et al (2013), Heikkilä et al (2012) and Espeland
531 et al (2018) were virtually identical but Condamine et al (2018) found clearly younger
532 ages. For the crown age of Riodinidae, there are also great discrepancies among
533 studies. Our core analysis (70.9, 95%, CI=57.2–85.2) gave identical results to
534 Heikkilä et al (2012) and Espeland et al (2018). Espeland et al (2015), in a study
535 focusing specifically on Riodinidae found about 10 million-year-older ages and
536 constitute the oldest estimate. Wahlberg et al (2013), however, found a much younger
537 estimate, about 20 Ma younger, in line with a recent study by Seraphim et al (2018)

specifically dedicated to the Riodinidae. For Nymphalidae, there is the greatest number of estimates, but they typically have relatively similar results. Our estimation (82.0, 95%, CI=68.1–98.3) was very close to that of Wahlberg et al (2013), Heikkilä et al (2012), Espeland et al (2018) and Condamine et al (2018, one clock) but about 12 million years younger than the study by Wahlberg et al (2009) who focused on Nymphalidae.

DISCUSSION

Fossils and Minimum Ages

In the core analysis we adopted a very conservative approach. This choice involves taking into account the uncertainty surrounding the information available for each calibration point, although at the expense of the amount of useful information available. For fossil constraints, this decision had two consequences. First, we calibrated the stem of the focal clade consisting of a fossil that was assigned by calibrating the divergence from its sister group, instead of the first divergence recorded in the phylogeny within the focal clade itself. Calibrating the crown age of the focal clade – meaning that we assume that the fossil is "nested" within the clade – may lead to an overestimation of the crown age. Such would be the case if lineages are undersampled at the root, or if extinction occurred, or if the fossil belongs to a lineage that actually diverges somewhere along the stem. Calibrating a deep node with the age of the fossil, which involves loss of some information, can help avoiding these problems. Second, we used uniform prior distributions bounded by the age of the fossil and the age of angiosperms. We considered that fossils provide only a minimum age for a node, a condition that is especially exacerbated by the exceptionally poor fossil record of Lepidoptera in general (Labandeira and Sepkoski,

562 1993) and Papilionoidea in particular (Sohn et al. 2015) when compared to the four
563 other major hyperdiverse insect lineages (Coleoptera, Hymenoptera, Diptera and
564 Hemiptera). Prior expectation on the age of the node cannot be modeled more
565 accurately without additional information. However, the marginal priors resulting
566 from the interactions among the different priors strongly differ from this assumption.

567

568 *Deep*- versus *Shallow-Level Calibrations*

569 Generally, favoring multiple calibrations placed at various positions in a tree instead
570 of a single or few calibrations, seem to produce more reliable estimates of molecular
571 clocks (Conroy & Van Tuinen 2003, Smith & Peterson 2002, Soltis et al 2002,
572 Duchêne et al 2014). Calibrations distributed across a tree may allow for a better
573 estimation of substitution rates and their pattern of variation among lineages
574 (Duchêne et al 2014), and consequently improve age estimates in cases of taxon
575 undersampling (Linder et al 2005).

576 Calibrations placed at deep levels in the tree are usually favored (Sauquet 2012, Hug
577 & Roger 2007) over calibrations at shallow levels for better capturing overall genetic
578 variation (Duchêne et al. 2014). Duchêne et al (2014) showed that using deep or
579 multiple calibrations particularly improves the estimation of substitution rates. Yet,
580 deep calibrations still tend to underestimate the mean substitution rate, especially
581 when substitution models are unable to correctly estimate the amount of "hidden"
582 substitutions along the deeper branches. Such underestimation can lead to an
583 overestimation of shallow node ages, referred to as "tree extension" by Phillips (2009).
584 For the butterflies, we investigated the consequences of using different subsets of
585 fossil calibrations according to their positions in the tree (deep *versus* shallow-level

586 calibrations), compared to the full set of fossil constraints. With a subset of fossils
587 placed only at deep levels in the phylogeny, we obtained results similar to the full set
588 of fossils in the core analysis, either at deep nodes or shallow nodes, indicating no tree
589 extension effect. This effect may also indicate that the shallow level calibration points
590 that are close to the tips are uninformative, and when included in the core analysis, do
591 not affect the timescale but clearly affected the priors (see below).

592 Alternatively, Duchêne et al (2014) showed that shallow-level calibrations can lead to
593 underestimation of the length of deep branches, thereby underestimating the timescale
594 and resulting in "tree compression" (Phillips 2009). We observed here a tree
595 compression effect since using only a subset of fossils placed close to the tips led to
596 the youngest estimates, including the credibility intervals. Also, we noticed in the core
597 analysis that nodes calibrated by *Protocoeliades* and *Vanessa* (two deep node
598 constraints) showed posterior distributions abutting against the minimum boundaries
599 defined by the age of the fossils, therefore preventing the tree (or at least these nodes)
600 to be younger in age.

601

602 *Host Plants and Maximum Ages*

603 For calibration points constrained by the age of the host-plant group, we considered
604 that only the crown of the focal clade could be assigned confidently to the host-plant
605 group, as the stem or part of the stem could be older than the host plant (the host-plant
606 shift would be happening somewhere along the stem). Support arises from molecular
607 biological and paleobiological evidence that the establishment of specialized insect–
608 herbivore associations can considerably postdate the origins of their hosts, as
609 illustrated in a Bayesian analysis of 100 species of leaf-mining *Phyllonorycter* moths

610 (Lepidoptera: Gracillariidae) and their dicot angiosperm hosts (Lopez-Vaamonde et al
611 2006). Relying on host-plant ages for calibrating a butterfly tree is questionable while
612 the timing of the divergence of angiosperms is still highly controversial (e.g.
613 Magallón et al 2015, Foster et al 2017). Therefore, first we calibrated our tree using
614 the oldest boundary of 95% CI of the stem age of a host-plant clade. This allowed us
615 to take into account the uncertainty surrounding the timing of the first appearance of
616 the host plant but consequently, it also relaxed the prior hypothesis for the calibrations.
617 Secondly, we compared two alternative timescales for the angiosperms: a
618 paleontological estimate, which infers an earlier Early Cretaceous origin of
619 angiosperms (Magallón et al 2015), and a molecular clock estimate that we extracted
620 from Foster et al (2017), which infers a stem age for angiosperms during the Early
621 Triassic, about 100 million years older. These two alternative scenarios affected the
622 size of the credibility intervals and the shape of the posterior distributions. For the
623 crown of Papilionoidea, the upper boundary of the 95%, CI was ~37 million years
624 older when using the molecular clock estimate. However, the shape of the distribution
625 was very asymmetrical, with a mode of the distribution very close to the core analysis
626 (101.0 Ma), showing that the estimation of the root still concentrated approximately at
627 the same ages. Using the hypothesis of an Early Triassic origin of angiosperms
628 implied very permissive priors toward old ages, which are most likely responsible for
629 the very wide credibility intervals and asymmetrical posterior distributions recovered
630 in the alternative analysis of using ages from Foster et al (2017). Therefore, it is
631 tempting to use the time-scale inferred using Magallón et al (2015)'s ages of
632 angiosperms, as it greatly narrows down the uncertainty surrounding butterfly ages,
633 and aligns more realistically with the fossil angiosperm record. However, as long as

634 there is no consensus on the timing of angiosperm diversification there is no reason to
635 favor one or the other hypothesis for an angiosperm origin timeline.

636 Alternatively, we also removed these maximum ages and focused only on the
637 information provided in the vetted list of fossils. Uniform priors can hardly be used
638 without a maximum age, so in this case we used lognormal priors. We found
639 credibility intervals narrower than the core analysis, while simply relaxing the host-
640 plant ages provided by Foster et al (2017) gave wider credibility intervals. This
641 strongly suggests that changing the shape of priors rather than removing maximum
642 constraints influenced the credibility intervals of the node ages.

643 *Priors and Posterior Distributions*

644 We compared the marginal priors to the posterior distributions for different analyses
645 of the root of Papilionoidea and for the different calibration points in the core analysis.
646 We found several calibration points showing a substantial shift of posterior
647 distribution. This indicates that our age estimates are not entirely driven by the set of
648 constraints, but instead the molecular dataset brings additional information about the
649 age of the calibrated nodes. An interesting pattern we found in the core analysis is the
650 consistent trend of posterior distributions of the shallow-level calibrated nodes to shift
651 toward older ages than the priors. Meanwhile, some deep-level node calibrations
652 shifted toward younger ages than the prior but most of them largely overlapped with
653 their prior distribution. Consequently, posterior estimates tend to contract the middle
654 part of tree compared to the prior estimates.

655 There are at least three reasons for the anomalous gap between the earliest fossil
656 papilionoid occurring at 55.6 Ma and its corresponding Bayesian median age of 110
657 Ma, that represents a doubling of the lineage duration. First, it long has been known

658 that the lepidopteran fossil record is extremely poor when compared to the far more
659 densely and abundantly occurring fossils of the four other hyperdiverse, major insect
660 lineages of Hemiptera, Coleoptera, Diptera and Hymenoptera (Labandeira and
661 Sepkoski, 1993). Second, particularly large-bodied apoditrysians such as
662 Papilionoidea, have even a poorer fossil record than other Lepidoptera in general,
663 particularly as they bear a fragile body habitus not amenable to preservation.
664 Additionally, as external feeders papilionoids lack a distinctive, identifiable, parallel
665 trace-fossil record such as leaf mines, galls and cases (Sohn et al 2015). Third, there
666 are very few productive terrestrial compression or amber deposits spanning the Upper
667 Cretaceous, from 100 Ma to the Cretaceous–Paleogene boundary of 66.0 Ma, and the
668 Paleogene Period interval from 66.0 Ma to the earliest papilionoid fossil of 55.6 Ma is
669 equally depauperate (Labandeira, 2014; Sohn et al., 2015). Some of these deposits
670 have recorded very rare small moth fossils, but to date no papilionoid, or for that
671 matter, other large lepidopteran taxa such as saturniids or pyraloids have been found.

672 The root of the tree was only calibrated with the oldest fossil in our dataset, a 55.6
673 million-year-old papilionoid, and the crown age of the angiosperms. However, the
674 prior distribution for the root in the core analysis clearly excluded an origin of
675 butterflies close to 55.6 Ma, but rather a distribution centered on a median of 110 and
676 a range of between 86.4 and 136.2 Ma. The posterior distribution for the root in the
677 core analysis largely overlapped with the prior. However, when we used alternative
678 ages for the angiosperms (older ages), the marginal prior for the root shifted to
679 substantially older ages. Nevertheless, the posterior distribution showed a significant
680 shift toward younger ages, albeit highly skewed, and toward ages similar to the core
681 analysis. This suggests that our estimate of the root age in the core analysis is not

simply driven by our set of priors, even if we do not actually observe a shift between marginal prior and posterior distributions.

We observed differences in prior and posterior distributions at the root when considering only subsets of fossils. When using only the subset of deep-level fossils, the marginal prior for the root showed very little difference from the core analysis prior and the posterior distributions completely overlapped. When using the subset of shallow-level fossils the marginal prior remained similar to the core analysis but the posterior distribution showed a substantial shift toward younger ages, yielding the youngest estimation of the age of Papilionoidea among all our analyses. As such, it seems that the choice of fossils did not change the prior estimation of the root, but the posterior distribution was largely influenced by deep-level fossils. As we suggested earlier, shallow-level fossils may be overestimating the mean substitution rate across the tree, and therefore underestimating the time scale, while the implementation of deep-level fossils seems to be correcting for this.

Timescale of Butterflies Revisited

We propose a new estimate for the timing of diversification of butterflies, based on an unprecedented set of fossil and host-plant calibrations. We estimated the origin of butterflies between 89.5 and 129.5 Ma, the median of this posterior distribution is 107.6 Ma, which corresponds to latest Early Cretaceous. The result of our core analysis for the root is very close to previous estimates by Wahlberg et al. (2013) and Heikkilä et al (2012). In comparisons of alternative analyses, the prior and posterior distributions showed that this result is robust to almost all the choices made throughout the core analysis and that our molecular dataset contains significant information in addition to the time constraints. This estimation means that there is a

52 million-year-long gap between the oldest known butterfly fossil and the molecular clock estimate. Interestingly, with more than 300 genes, Espeland et al (2018) found ages very similar to ours, suggesting that our estimates are not due to the lack of information contained in our molecular dataset to estimate the molecular clock. Alternatively, the fossil record for butterflies is so sparse that an intervening fossil gap is highly likely. Additionally, the fossil *Protocoeliades kristenseni*, which is 55.6 Ma can be assigned confidently to the crown of the family Hesperiidae and the stem of Coeliadinae, which is well within the Papilionoidea clade. For angiosperms, a very rich fossil record is available compared to butterflies (e.g., Magallón et al (2015), which used 137 fossils to calibrate a phylogeny of angiosperms), rendering the absence of angiosperms, either as pollen or macrofossils, that are older than 136 Ma much more puzzling.

FUNDING

NW acknowledges funding form the Swedish Research Council (Grant No. 2015-04441) and from the Department of Biology, Lund University. AVLF thanks CNPq (grant 303834/2015-3), National Science Foundation (DEB-1256742) and FAPESP (grant 2011/50225-3). This publication is part of the RedeLep (Rede Nacional de Pesquisa e Conservação de Lepidópteros) SISBIOTABrasil/CNPq (563332/2010-7). MH gratefully acknowledges funding from a Peter Buck Postdoctoral Stipend, Smithsonian Institution National Museum of Natural History. JCS is receiving supports from the Research Under Protection program (NRF-2017R1D1A2B05028793), funded by the National Research Foundation of Korea.

ACKNOWLEDGMENTS

This is contribution 366 of the Evolution of Terrestrial Ecosystems consortium at the National Museum of Natural History, in Washington D.C.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository: doi:10.5061/dryad.fb88292.

LITERATURE CITED

Ackery P.R. 1988. Hostplants and classification: a review of nymphalid butterflies. Biol J Linnean Soc, 33(2): 95-203.

Beccaloni G.W., Viloria A.L., Hall S.K., Robinson G.S. 2008. Catalogue of the hostplants of the Neotropical butterflies. Sociedad Entomológica Aragonesa (SEA), Zaragoza, Spain.

Braby M.F., Vila R., Pierce N.E. 2006. Molecular phylogeny and systematics of the Pieridae (Lepidoptera: Papilionoidea): higher classification and biogeography. Zool. J. Linn. Soc. 147:239–275.

Brenner G.J. 1996. Evidence for the earliest stage of angiosperm pollen evolution: a paleoequatorial section from Israel. In: Taylor, D.W., Hickey, L.J., editors, Flowering plant origin, evolution & phylogeny. Boston, MA: Springer. P. 91–115.

Brown J.W., Smith S.A. 2017. The past sure is tense: On interpreting phylogenetic divergence time estimates. Syst. Biol., 66, doi: 10.1093/sysbio/syx074

751 Chazot N., Willmott K.R., Condamine F.L., De - Silva D.L., Freitas A.V., Lamas G.,
752 Morlon H., Giraldo C.E., Jiggins C.D., Joron M., Mallet J., Elias M. 2016. Into the
753 Andes: multiple independent colonizations drive montane diversity in the Neotropical
754 clearwing butterflies Godyridina. Molecular Ecology. 25:5765-5784.

755 Cockerell T.D.A. 1907. A fossil butterfly of the genus *Chlorippe*. Canad. Entomol.,
756 39:361–363,

757 Condamine F.L., Sperling F., Wahlberg N., Rasplus J.-Y., Kergoat G.J. 2012. What
758 causes latitudinal gradients in species diversity? Evolutionary processes and
759 ecological constraints on swallowtail biodiversity. Ecology Letters. 15:267-77.

760 Condamine F.L., Nagalingum N., Marshall. C., Morlon H. 2015. Origin and
761 diversification of living cycads: A cautionary tale on the impact of the branching
762 process prior in Bayesian molecular dating. BMC evolutionary biology. 15:65.

763 Condamine F.L., Nabholz B., Clamens A.-L., Dupuis J.R., Sperling F.A. 2018.
764 Mitochondrial Phylogenomics, the origin of swallowtail butterflies, and the impact of
765 the number of clocks in Bayesian molecular dating. Syst. Entomol 43:460–480.

766 Conroy C.J., van Tuinen M. 2003. Extracting time from phylogenies: positive
767 interplay between fossil and genetic data. J. Mammal. 84:444–455.

768 de Jong R. 2016. Reconstructing a 55-million-year-old butterfly (Lepidoptera:
769 Hesperiidae). Eur. J. Entomol. 113:423–428.

770 de Jong R. 2017. Fossil butterflies, calibration points and the molecular clock
771 (Lepidoptera: Papilionoidea). Zootaxa 4270:1–63.

772 dos Reis M., Yang Z., 2013. The unbearable uncertainity of Bayesian divergence time
773 estimation. J. Syst. Evol. 51:30–43.

774 Drummond A. J., Ho S. Y., Phillips M. J., Rambaut A. 2006. Relaxed phylogenetics
775 and dating with confidence. *PLoS biology*, 4(5), e88.

776 Drummond A.J., Suchard M.A., Xie D., Rambaut A. 2012. Bayesian phylogenetics
777 with BEAUTi and BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

778 Duchêne S., Lanfear R., Ho S.Y.W 2014. The impact of calibration and clock-model
779 choice on molecular estimates of divergence times. Mol. Phylogen. Evol., 78:277–289.

780 Durden C.J., Rose H. 1978. Butterflies from the Middle Eocene: the earliest
781 occurrence of fossil Papilionoidea (Lepidoptera). The Pearce-Sellards Series of the
782 Texas MemorialMuseum. 29:1–25.

783 Ehrlich P.R., Raven P.H. 1964. Butterflies and plants: a study in coevolution.
784 Evolution. 18 (4). 586-608.

785 Espeland M., Breinholt J., Wilmott K.R., Warren A.D., Vila R., Toussaint E.F.A.,
786 Maunsell S.A., Aduse-Poku K., Tatavera G., Eastwood R., Jarzyna M.A., Guralnick
787 R., Lohman D.J., Pierce N.E., Kawahara A.V. 2018. A comprehensive and dated
788 phylogenomic analysis of butterflies. Curr. Biol. 28: 1–9 (doi:
789 10.1016/j.cub.2018.01.061).

790 Foster C.S.P., Sauquet H., van der Merwe M., McPherson H., Rosette M., Ho S.Y.W.
791 2017. Evaluating the impact of genomic data and priors on Bayesian estimates of the
792 angiosperm evolutionary timescale. Syst. Biol., 66:338–351.

793 Garzón-Orduña I.J., Silva-Brandão K.L., Wilmott K.R., Freitas A.V.L., Brower
794 A.V.Z. 2015. Incompatible ages for clearwing butterflies based on alternative
795 secondary calibrations. Syst. Biol., 64:752–767.

796 Hall J.P.W., Robins R.K., Harvey D.J. 2004. Extinction and biogeography in the
797 Caribbean: new evidence from a fossil riodinid butterfly in Dominican amber. Proc. R.
798 Soc. Lond. B, 271:797–801.

799 Heer O. 1849. Die Insektenfauna der Tertiärgebilde von Oeningen und von Radoboj
800 in Croatien. Vol. 2. Wilhelm Engelsmann, Leipzig, 264 pp.

801 Heikkilä M., Kaila L., Mutanen M., Peña C., Wahlberg N. 2012. Cretaceous origin
802 and repeated Tertiary diversification of the redefined butterflies. Proc. R. Soc. Lond.
803 B, 279:1093–1099.

804 Hug L.A., Roger A.J. 2007. The impact of fossils and taxon sampling on ancient
805 molecular dating analyses. Mol. Biol. Evol. 24:1889–1897.

806 International Commission on Stratigraphy. 2012. International Chronostratigraphic
807 Chart: http://www.stratigraphy.org/ICSchart/ChronostratChart2012.pdf

808 Janz, N. Nylin S. 1998. Butterflies and plants: a phylogenetic study. Evolution, 52(2):
809 486-502.

810 Kozak K.M., Wahlberg N., Neild A.F.E., Dasmahapatra K.J.K., Jiggins C.D. 2015.
811 Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius*
812 butterflies. Syst. Biol., 64:505–524,

813 Labandeira C.C. 2014. Amber. In: Laflamme M., Schiffbauer J.D., Darroch S.A.F.,
814 editors, Reading and writing of the fossil record: Preservational pathways to
815 exceptional fossilization. Paleontol. Soc. Pap., 20:163–216.

816 Labandeira C.C. Sepkoski J.J., Jr. 1993. Insect diversity in the fossil record. Science,
817 261:310–315.

818 Lanfear R., Calcott B., Ho S.Y.W., Guindon S. 2012. PartitionFinder: combined
819 selection of partitioning schemes and substitution models for phylogenetic analysis.
820 Mol. Biol. Evol. 29:1695–1701.

821 Linder H.P., Hardy C.R., Rutschmann F. 2005. Taxon sampling effects in molecular
822 clock dating: An example from the African Restionaceae. Mol. Phylogen. Evol.
823 35:569–582.

824 Lopez-Vaamonde C., Wikström N., Labandeira C., Godfray H.C.J., Goodman S.J.,
825 Cook J.M. 2006. Fossil-calibrated molecular phylogenies reveal that leaf-mining
826 moths radiated millions of years after their host plants. J. Evol. Biol., 19:1314–1326,

827 Magallón S. Gómez-Acevedo S., Sánchez-Reyes L.L., Hernández-Hernández T. 2015.
828 A metacalibrated time-tree documents the early rise of flowering plant phylogenetic
829 diversity. New Phytol., 207:437–453.

830 Martins-Neto R.G., Kucera-Santos J.C., Vieira F.R. de M., Fragoso L.M. de C. 1993.
831 Nova espécie de borboleta (Lepidoptera: Nymphalidae: Satyrinae) da Formação
832 Tremembé, Oligoceno do Estado de São Paulo. Acta Geol. Leopold., 37:5–16.

833 Matos-Maravi P.F., Peña C., Wilmott K.R., Freitas A.V.L., Wahlberg N. 2013.
834 Systematics and evolutionary history of butterflies in the "*Taygetis* clade"

(Nymphalidae: Satyrinae: Euptychiina): towards a better understanding of Neotropical biogeography. Mol. Phylogenet. Evol. 66:54–68.

Miller J.Y. Brown F.M. 1989. A new Oligocene fossil butterfly, *Vanessa* †*amerindica* (Lepidoptera: Nymphalidae), from the Florissant Formation, Colorado. Bull. Allyn Mus., 126:1–9.

Nel A., Nel J., Balme C. 1993. Un nouveau Lépidoptère Satiyrinae fossile de l'Oligocene du Sud-Est de la France (Insecta, Lepidoptera, Nymphalidae). Linn. Belg., 14:20–36.

Nylin S., Janz, N. 1999. Ecology and evolution of host plant range: butterflies as a model group. In: Olff H., Brown V.K., Drent R.H. editors. Herbivores: between plants and predators. Blackwell Science Ltd, Oxford. p 31–54.

Paradis E., Claude J., Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289-290.

Parham J.F., Donoghue P., Bell C.J., Calway T.D., Head J.J., Holroyd P.A., Inoue J.G., Irmis R.B., Joyce W.G., Ksepka D.T., Patané J.S.L., Smith N.D., Tarver J.E., van Tuinen M., Yang Z., Angielczyk K.D., Greenwood J.M., Hipsley C.A., Jacobs L., Mackovicky P.J., Miller J., Smith K.T., Theodor J.M., Warnock R.C.M., Benton M.J. 2012. Best practices for justifying fossil calibrations. Syst. Biol., 61:346–359.

Peña C., Malm T. 2012. VoSeq: A voucher and DNA sequence web application. PLoS ONE, 7(6):e39071.

855 Peña C., Nylin S., Wahlberg N. 2011. The radiation of Satyrini butterflies
856 (Nymphalidae: Satyrinae): a challenge for phylogenetic methods. Zool J Linnean Soc,
857 161(1), 64-87.

858 Peñalver E., and Grimaldi D.A. 2006. New data on Miocene butterflies in Dominican
859 Amber (Lepidoptera: Riodinidae and Nymphalidae) with the description of a new
860 nymphalid. Am. Mus. Novit. 3591:1–17.

861 Phillips M.J. 2009. Branch-length estimation bias misleads molecular dating for a
862 vertebrate mitochondrial phylogeny. Gene, 441:132–140.

863 R Development Core Team 2008. R: A language and environment for statistical
864 computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-
865 900051-07-0, URL http://www.R-project.org.

866 Rambaut A., Suchard M. A., Xie D., & Drummond A. J. 2014. Tracer v1.6, Available
867 from http://tree.bio.ed.ac.uk/software/tracer/.

868 Rannala B. and Yang Z. 2007. Inferring speciation times under an episodic molecular
869 clock. Syst. Biol., 56:453–466.

870 Rebel H. 1898. Fossile Lepidopteren aus der Miocän-Formation von Gabbro.
871 Sitzungsber. Akad. Wiss. Wien, 107:731–745.

872 Sahoo R. K., Warren A. D., Collins S. C., Kodandaramaiah U. 2017. Hostplant
873 change and paleoclimatic events explain diversification shifts in skipper butterflies
874 (Family: Hesperiidae). BMC Evolutionary Biology, 17: 174. Doi: 10.1186/s12862-
875 017-1016-x

876 Sahoo R.K., Warren A.D., Wahlberg N., Brower A.V.Z., Lukhtanov V.A.,
877 Kodandaramaiah U. 2016. Ten genes and two topologies: an exploration of higher
878 relationships in skipper butterflies (Hesperiidae). Peer J, 4:e2653, doi:
879 10.7717/peerj.2653

880 Sauquet H., Ho S.Y.W., Gandolfo M.A., Jordan G.I., Wilf P., Cantrill D.J., Bayly
881 M.J., Bromham L., Brown G.K., Carpenter R.J., Lee D.M., Murphy D.J., Sniderman
882 J.M.K., Udovice F. 2012. Testing the impact of calibration on molecular divergence
883 times using a fossil-rich group: The case of *Nothofagus* (Fagales). Syst. Biol.,
884 61:289–313.

885 Scudder S.H. 1875. Fossil butterflies. Mem. Am. Assoc. Adv. Sci. 1:1–99.

886 Scudder S.H. 1889. The fossil butterflies of Florissant. United States Geological
887 Survey, 8th Annual Report, pp. 439–472.

888 Seraphim N., Kaminsky L.A., DeVries P.J., Penz C., Callaghan C., Wahlberg N.,
889 Silva-Brandão K.L., Freitas A.V.L. 2018. Molecular phylogeny and higher
890 systematics of the metalmark butterflies (Lepidoptera: Riodinidae). Syst. Entomol.
891 43:407–425.

892 Smith A.B., Peterson K.J. 2002. Dating the time of origin of major clades: Molecular
893 clocks and the fossil record. Annu. Rev. Earth Planet. Sci. 30:65–88.

894 Sohn J.C., Labandeira C., Davis D., Mitter C. 2012. An annotated catalog of fossil
895 and subfossil Lepidoptera (Insecta: Holometabola) of the world. Zootaxa 3286:1–132.

Sohn J.C., Labandeira C.C., Davis D.R. 2015. The fossil record and taphonomy of butterflies and moths (Insecta, Lepidoptera) and implications for evolutionary diversity and divergence-time estimates. BMC Evol. Biol. 15:12.

Soltis D.E., Soltis P.S., Zanis M.J. 2002. Phylogeny of seed plants based on evidence from eight genes. Am. J. Bot. 89:1670–1681.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Toussaint E.F.A., Balke M. 2016. Historical biogeography of *Polyura* butterflies in the oriental Palaeotropics: trans-archipelagic routes and South Pacific island hopping. J. Biogeogr. 43:1560–1572.

Wahlberg N., Leneveu J., Kodandaramaiah U., Peña C., Nylin S., Freitas A.V.L., Brower A.V.Z. 2009. Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. Proc. R. Soc. Lond. B, 276:4295–4302.

Wahlberg N., Wheat C.W., Peña C. 2013. Timing and patterns in the taxonomic diversification of Lepidoptera (butterflies and moths). PLoS ONE, 8(11):e80875.

Walker J.D., Geissman J.W., Bowring S.A., Babcock L.E. 2013. The Geological Society of America geologic time scale. GSA Bull., 125:259–272.

Warnock R.C.M., Yang Z., Donoghue P.C.J. 2012. Exploring uncertainty in the calibration of the molecular clock. Biol. Lett., 9:156–159.

Warnock R.C.M., Parham J.F., Joyce W.G., Lyson T.R., Donoghue,´ P.C.J. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. Proc. R. Soc. Lond. B, 282:20141013.

Yang Z, Rannala B. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol., 23:212–226.

TABLE 1. (a) Fossil calibration points used to calibrate the tree as a minimum age for the *Clade calibrated*. Unless stated otherwise, the fossil calibrations were placed at the stem of the clade calibrated. *Lower* and *upper* values indicate the prior truncation for both the uniform and exponential priors. The 140 Ma year upper truncation corresponds to the maximum age of Angiosperms from Magallón et al 2015. A different upper truncation value results from a fossil prior interacting with a host-plant prior placed at the same node or a shallow node. *Mean* and *offset* are parameter values for the exponential prior distribution. (b) Parameter values used for the analysis using fossil information only, modeled using lognormal prior distributions. (c) Host-plant clades used to calibrate the tree as a maximum age for the *Clade calibrated* node. Host-plant calibrations were placed at the crown of the clade calibrated. Ages from both Magallón et al (2015) and Foster et al (2017) are indicated.

a)

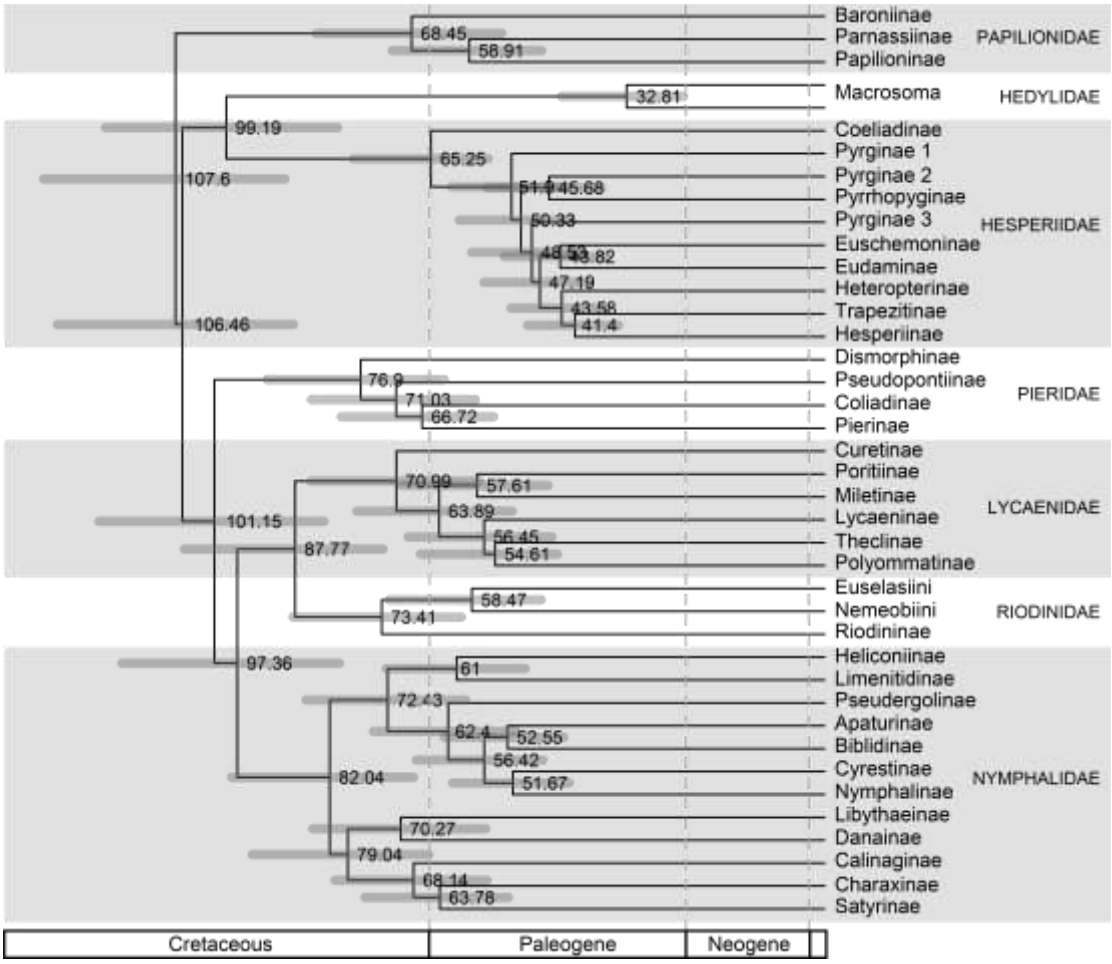
Fossils	Clade calibrated	lower	upper	mean	offset
Doritites bosniaskii Rebel, 1898	Papilionidae: Parnassiinae: Luehdorfiini	5.3	140	25	5.3
Dynamine alexae Peñalver & Grimaldi, 2006	Nymphalidae: Biblidinae: *Dynamine*	15.9	89	20	15.9
Lethe corbieri Nel, Nel & Balme, 1993	Nymphalidae: Satyrinae: Satyrini	28.3	65	25	28.3
Mylothrites pluto Heer, 1849	Pieridae: Coliadinae+Pierinae	15.9	100	50	15.9
Neorinella garciae Martins-Neto et al., 1993	Crown of Amathusiini	23.0	65	20	23.0
Pamphilites abdita Scudder, 1875	Hesperiidae: Hesperiinae	23.0	140	30	23.0
Prolibythea vagabunda Scudder, 1889	Nymphalidae: Libytheinae	33.8	140	40.0	33.8
Protocoeliades kristenseni de Jong, 2016	Hesperiidae: Coeliadinae	55.6	140	35	55.6
Thaites ruminiana Scudder, 1875	Papilionidae: Parnassiinae: Parnassiini	23.0	140	25	23.0
Theope sp	Riodinidae: Riodininae: Nymphidiini: *Theope*	15.9	140	25	15.9
Voltinia dramba Hall, Robinson & Harvey, 2004	Riodinidae: Riodininae: Eurybiini: *Voltinia*	15.9	140	30	15.9
Vanessa amerindica Miller & Brown, 1989	Nymphalidae: Nymphalinae: Nymphalini	33.8	140	30	33.8
Doxocopa wilmattae Cockerell, 1907	Nymphalidae: Nymphalinae+Biblidinae+ Limenitidinae+Apaturinae		Not used		
Praepapilio colorado Durden & Rose, 1978	Papilionidae		Not used		

b)

Fossils	mean	stdev	offset
Doritites bosniaskii Rebel, 1898	25	0.9	5.3
Dynamine alexae Peñalver & Grimaldi, 2006	30	0.9	15.9
Lethe corbieri Nel, Nel & Balme, 1993	40	1	28.3
Mylothrites pluto Heer, 1849	15.9	100	50
Neorinella garciae Martins-Neto et al., 1993	30	1	23.0
Pamphilites abdita Scudder, 1875	30	1	23.8
Prolibythea vagabunda Scudder, 1889	50	0.8	33.8
Protocoeliades kristenseni de Jong, 2016	70	0.8	35.0
Thaites ruminiana Scudder, 1875	40	1	23.0
Theope sp	30	1	15.9
Voltinia dramba Hall, Robinson & Harvey, 2004	30	1	15.9
Vanessa amerindica Miller & Brown, 1989	45	1	28.2
Doxocopa wilmattae Cockerell, 1907		Not used	
Praepapilio colorado Durden & Rose, 1978		Not used	

c)

Host-plant clade	Clade calibrated	Magallón et al. 2015	Foster et al. 2017
Angiospermae	root	140	252
Poaceae	Hesperiidae: Hesperiinae	65	112
Poaceae	Nymphalidae: Satyrinae	65	112
Fabaceae	Pieridae	100	123
Brassicaceae	Pieridae: Pierinae	103	97
Rubiaceae	Riodinidae: Leucochimona+Mesophtalma+ Mesosemia+Perophthalma +Semomesia	87	85
Apocynaceae	Nymphalidae: Danainae	69	85
Solanaceae	Nymphalidae: Ithomiini	87	68
Euphorbiaceae	Nymphalidae: Biblidinae	89	104
Sapindaceae	Nymphalidae: Biblidinae: Epiphilini+ Callicorini	87	91

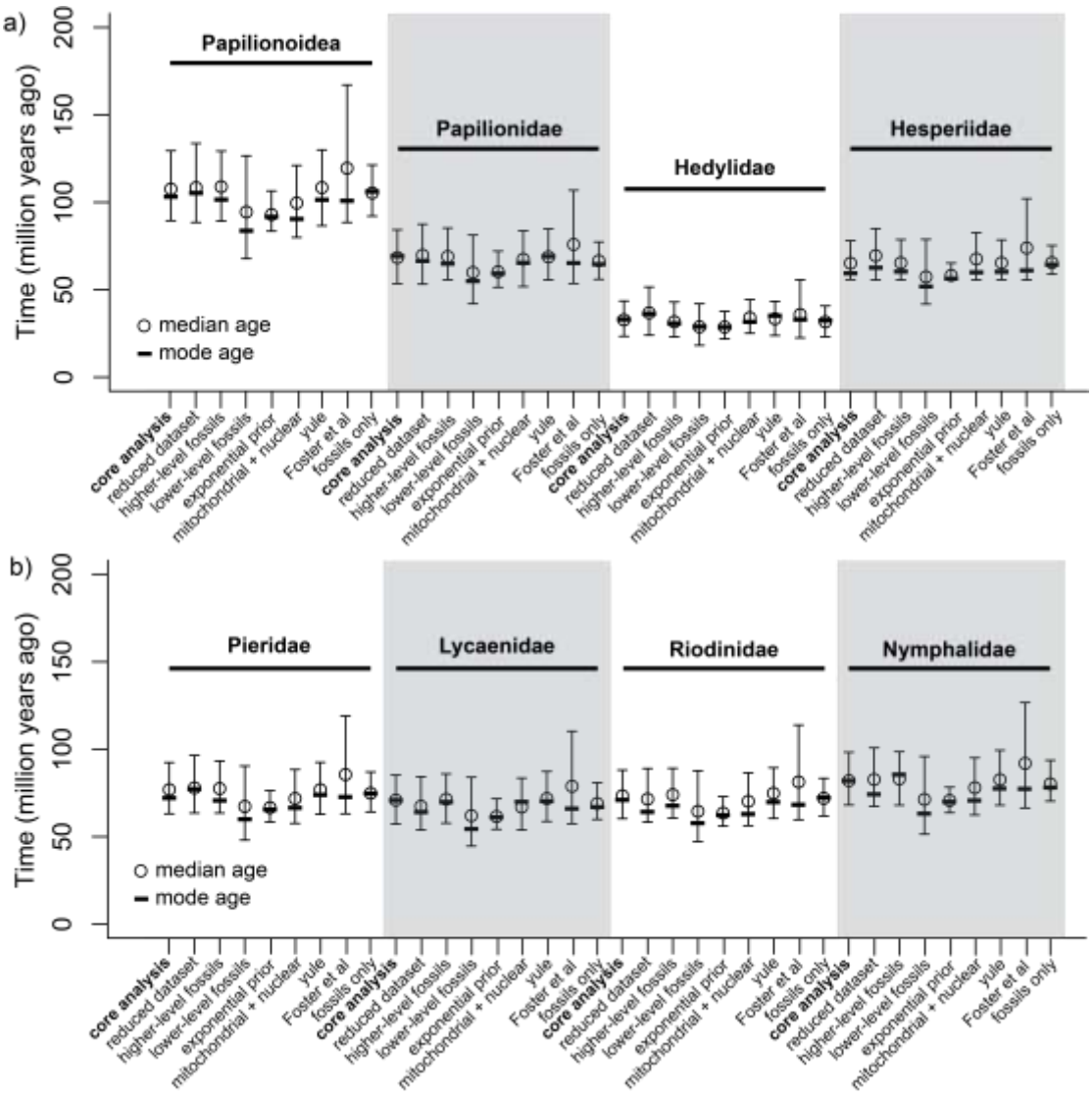


995

996 FIGURE 1. Time-calibrated tree obtained from the core analysis. Only the relationships
997 and age estimates among the subfamilies of Papilionoidea are shown here. The
998 complete tree, including median node ages, credibility intervals, and the positions of
999 fossil and host-plant calibration points are shown in Supplementary Material S12.
1000 Age estimates are indicated at the nodes (Ma). Node bars represent the 95%
1001 credibility intervals.

1002

1003



1004

1005 FIGURE 2. Comparison of node age estimates for the root of Papilionoidea and the

1006 seven families between the core analysis and the seven alternative analyses. Mode,

1007 median and 95% credibility interval are presented.

1008

1009

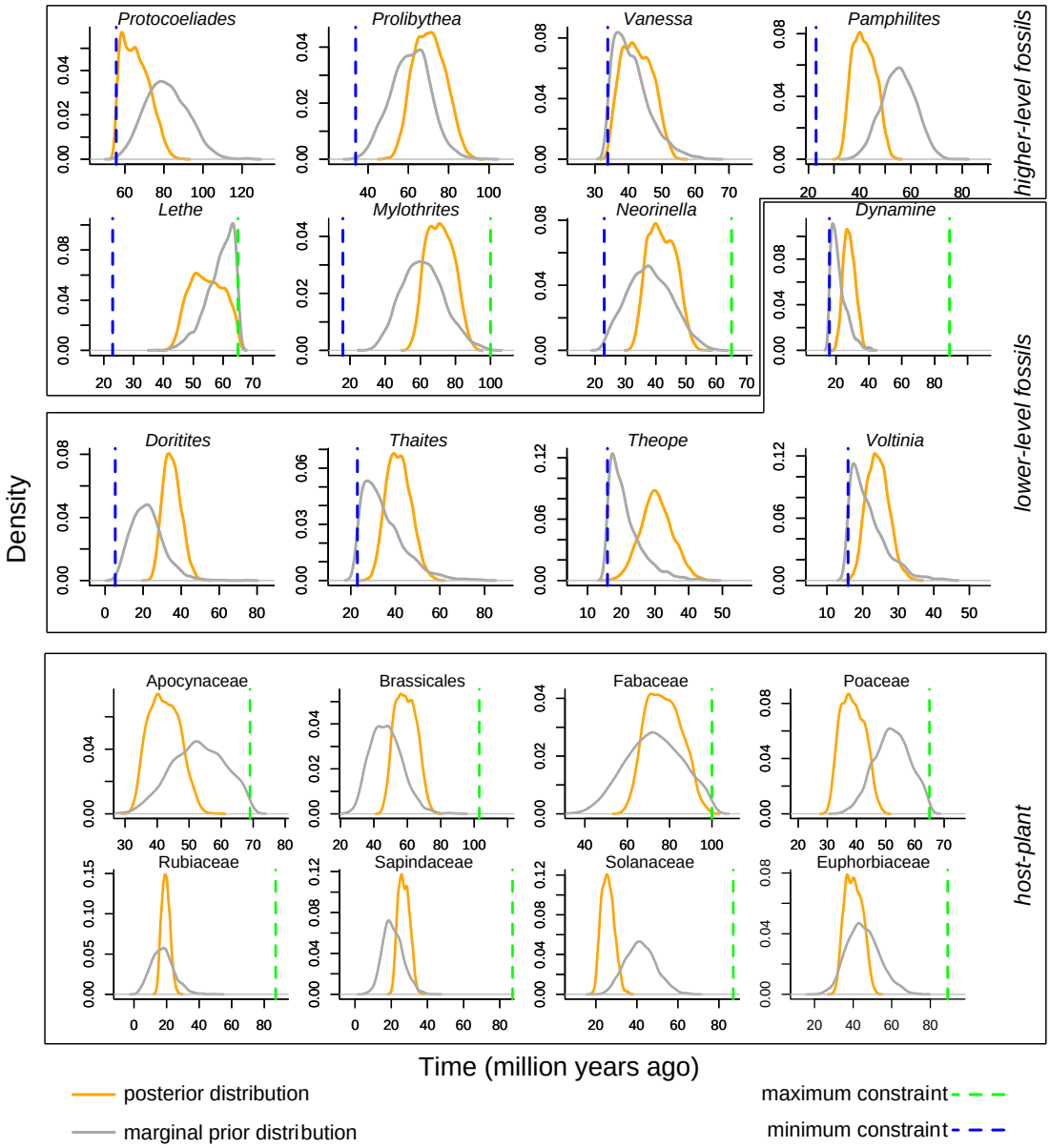
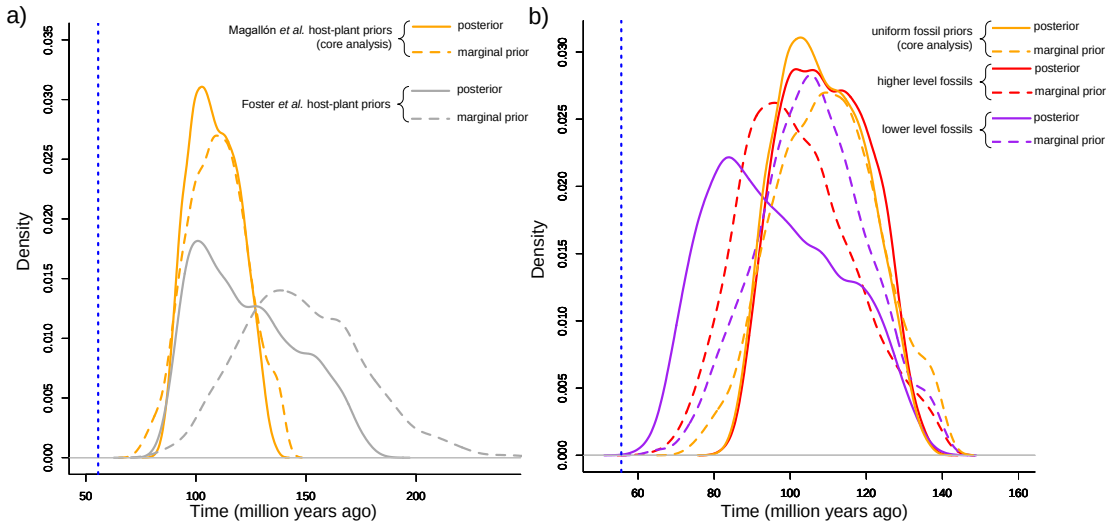


FIGURE 3. Marginal prior (grey) and posterior distributions (orange) for the nodes calibrated in the core analysis. Blue dashed lines represent minimum boundaries; green dashed lines represent maximum boundaries.



1017

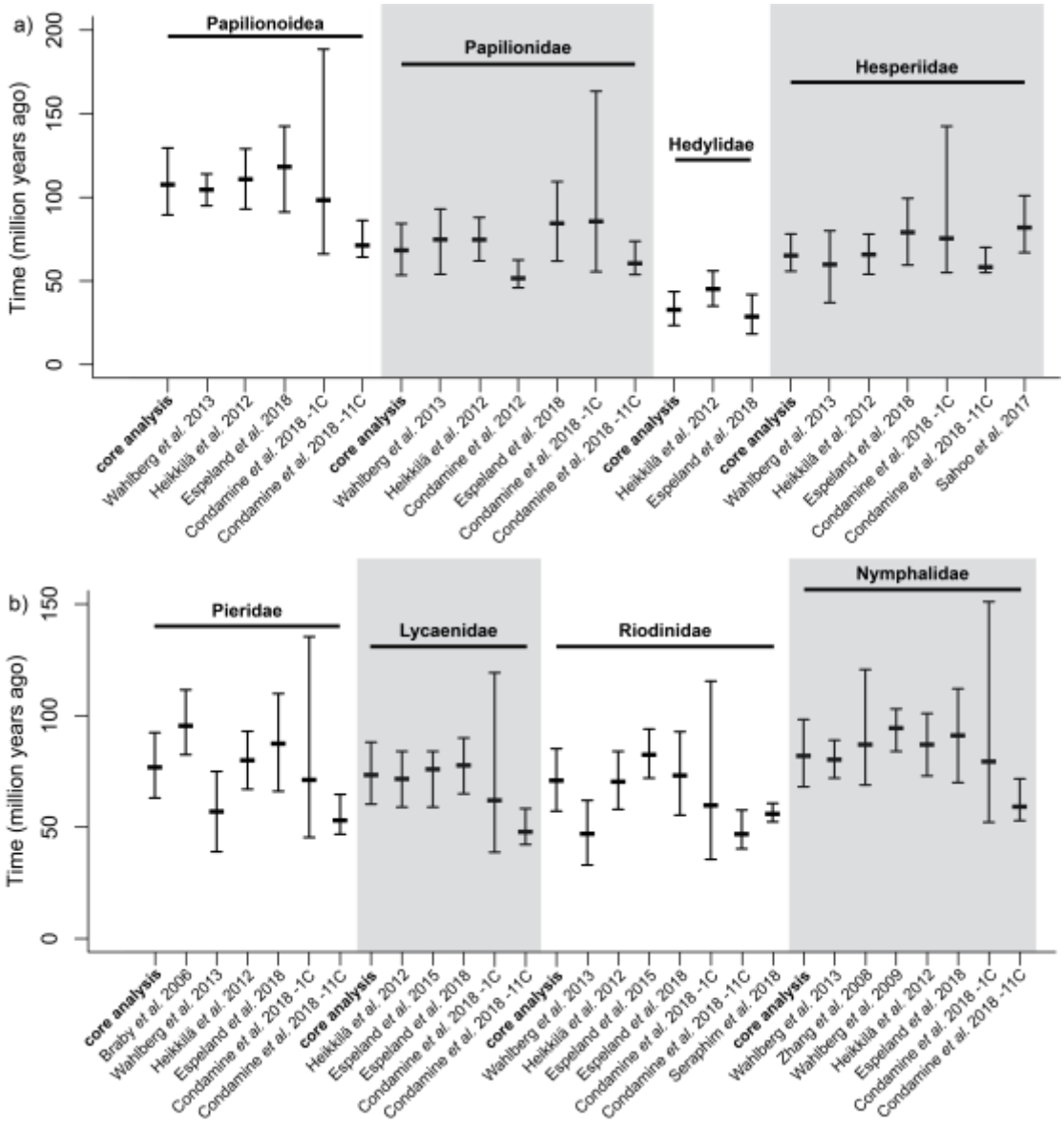
1018 FIGURE 4. Marginal prior and posterior distributions for the root age in the core

1019 analysis using either a) alternative host-plant ages or b) Alternative subsets of fossil

1020 calibrations.

1021

1022



1023

1024 FIGURE 5. Comparison of node age estimates for the root of Papilionoidea and the

1025 seven families between this study (core analysis) and estimates from previous studies.

1026 Mode and 95% CI for the core analysis are presented. For the other studies the values

1027 reported in the original study are used.

1028